



Single Viewpoint Stereoscopic Sensor Calibration

Guillaume Caron, El Mustapha Mouaddib, Eric Marchand

► To cite this version:

Guillaume Caron, El Mustapha Mouaddib, Eric Marchand. Single Viewpoint Stereoscopic Sensor Calibration. Int. Symp. on Image/Video Communications over fixed and mobile networks, ISIVC-2010, 2010, Rabat, Morocco, Morocco. inria-00544783

HAL Id: inria-00544783

<https://inria.hal.science/inria-00544783>

Submitted on 10 Dec 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Single Viewpoint Stereoscopic Sensor Calibration

Guillaume Caron and El Mustapha Mouaddib
 Université de Picardie Jules Verne
 MIS Laboratory
 Amiens, France
 Email: {guillaume.caron, mouaddib}@u-picardie.fr

Eric Marchand
 Université de Rennes 1
 IRISA-INRIA Lagadic
 Rennes, France
 Email: eric.marchand@irisa.fr

Abstract—Calibration is a classical issue in computer vision. This work presents a camera calibration approach for single viewpoint cameras which can be modeled by the unified spherical projection. The paper extends the method of perspective camera calibration using virtual visual servoing, to central cameras and particularly to omnidirectional cameras. The contribution is the development of simultaneous intrinsic and extrinsic calibration of N central cameras stereo rig.

Index Terms—Omnidirectional vision, stereovision, calibration.

I. INTRODUCTION

A. Overview

Single viewpoint cameras are a family of sensors of which viewlines intersect themselves at a unique point. Perspective or catadioptric cameras with some optics and mirror type [1] are central.

Stereovision generally uses several cameras on a rig. In central omnidirectional vision, a few systems exist using two catadioptric cameras [2] or one camera and two or more mirrors [3], [4] (Fig. 1).

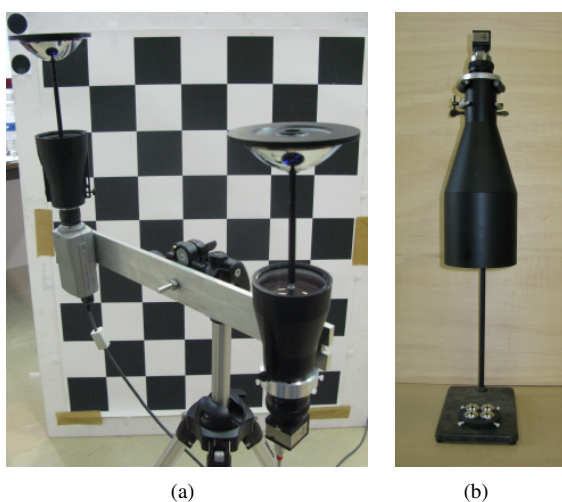


Fig. 1. Stereo catadioptric sensors. (a) a classical way to make an omnidirectional stereo rig. It can also be vertical leading to different properties [2]. (b) Omnidirectional stereovision is achieved with only one camera and four mirrors [4].

Central cameras can be modeled by the unified spherical projection model [5]. The calibration of such a parametric model is a numerical process allowing to find parameter values

of the model. This paper proposes a new approach to estimate projection parameters of a central camera and particularly of a stereo rig composed by N central cameras.

Calibration is necessary to recover metric information from images. This is a common problem as well in perspective vision [7], [6] as in omnidirectional vision [8], [9], [10]. Marchand *et al.* [6] proposed a virtual visual servoing (VVS) based perspective camera calibration. VVS [11], [12] is a non-linear optimization technique useful for pose computation and extendable to calibration, receiving the wide knowledge of visual servoing.

Works about calibration of stereo rigs generally tackle perspective cameras [7] or are specific to a sensor. This paper extends the VVS based central stereoscopic pose estimation method [13] to calibration of N cameras using points detected in images as corners of a known chessboard.

After a recall of the unified projection model, VVS based calibration of central camera and then stereoscopic system are tackled. Finally, calibration results are presented and compared to an existing monocular method.

B. Central sensor modeling

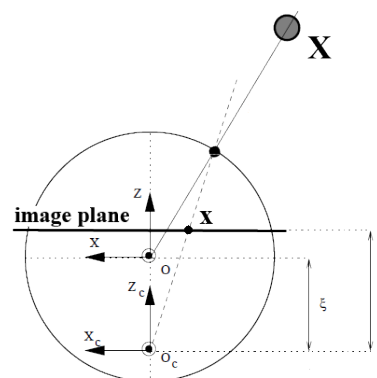


Fig. 2. Unified spherical projection model.

Following the unified spherical projection model for central cameras [5], a 3D point $X = (X \ Y \ Z)^T$ is first projected onto a unitary sphere, centered at $(0 \ 0 \ \xi)^T$ (Fig. 2). The obtained point is then perspectively projected on the image plane as $x = (x \ y \ 1)^T$, knowing intrinsic parameters

$\gamma = \{p_x, p_y, u_0, v_0, \xi\}$:

$$\mathbf{x} = pr_\xi(\mathbf{X}) \quad \text{with} \quad \begin{cases} x = \frac{X}{Z+\xi\rho} \\ y = \frac{Y}{Z+\xi\rho} \end{cases}. \quad (1)$$

where $\rho = \sqrt{X^2 + Y^2 + Z^2}$. $\mathbf{x} = (x \ y \ 1)^T$ is the point on the normalized image plane and the pixelic point $\mathbf{u} = (u \ v \ 1)^T$ is obtained by the relation $\mathbf{u} = \mathbf{K}\mathbf{x}$. \mathbf{K} is the intrinsic parameters matrix including p_x, p_y , the generalized focus and u_0, v_0 , the coordinates of the principal point. We also define $pr_\gamma(\mathbf{X}) = \mathbf{K}pr_\xi(\mathbf{X})$.

C. Stereovision

For a rig of N cameras, each camera is modeled by the unified model. Poses of $N-1$ cameras of the rig, modeled by homogeneous matrices ${}^{c_j}\mathbf{M}_{c_1}$, are defined w.r.t. the reference one:

$${}^{c_j}\mathbf{M}_{c_1} = \begin{pmatrix} \mathbf{R}_{3 \times 3} & \mathbf{t}_{3 \times 1} \\ 0_{1 \times 3} & 1 \end{pmatrix}. \quad (2)$$

II. CENTRAL SENSOR CALIBRATION

The pose computation problem is defined as a VVS issue. Usually, image based visual servoing aims to move a camera to a desired pose minimising errors between current image features and features of the image acquired at the desired pose. The VVS virtualizes the camera and starting from an initial pose, moves the virtual camera to make a perfect correspondence between the object forward projection, for its virtual pose, and the object in the real image.

The virtual camera is defined by its projection function $pr_\gamma()$ and its position ${}^c\mathbf{M}_o$. With \mathbf{r} , the pose vector, the method estimates the real pose minimizing the error Δ between detected points \mathbf{u}_i^* and current points obtained by forward projection for the current pose $\mathbf{u}_i(\mathbf{r})$:

$$\Delta = \sum_{i=1}^k (pr_\gamma({}^c\mathbf{M}_o, {}^o\mathbf{X}_i) - \mathbf{u}_i^*)^2. \quad (3)$$

The error to be regulated is, hence, $\mathbf{e} = \mathbf{u}(\mathbf{r}) - \mathbf{u}^*$. Imposing an exponential decrease of the error, the features motion is linked to the virtual camera by $\dot{\mathbf{e}} = -\lambda\mathbf{e}$, only depending of $\dot{\mathbf{u}}$ [6]:

$$\dot{\mathbf{u}} = \frac{\partial \mathbf{u}}{\partial \mathbf{r}} \frac{d\mathbf{r}}{dt} + \frac{\partial \mathbf{u}}{\partial \gamma} \frac{d\gamma}{dt} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{r}} \frac{d\mathbf{r}}{dt} + \frac{\partial \mathbf{u}}{\partial \gamma} \frac{d\gamma}{dt}. \quad (4)$$

$\mathbf{L}_\mathbf{x} = \frac{\partial \mathbf{x}}{\partial \mathbf{r}}$ is known as the pose interaction matrix related to a normalized omnidirectional image plane point \mathbf{x} [14] and other Jacobians are easy to compute.

The VVS control law for calibration using p images is then [6]:

$$(\mathbf{v}^1 \ \mathbf{v}^2 \ \dots \ \mathbf{v}^p \ \dot{\gamma})^T = -\lambda \mathbf{H}^+ \begin{pmatrix} \mathbf{u}^1 - \mathbf{u}^{*1} \\ \mathbf{u}^2 - \mathbf{u}^{*2} \\ \vdots \\ \mathbf{u}^p - \mathbf{u}^{*p} \end{pmatrix} \quad (5)$$

with \mathbf{v}^i , the camera pose velocity vector associated to image i , $\dot{\gamma}$ the time variation of intrinsic parameters (unique for the

images set) and \mathbf{u}^{*i} , the points set detected in image i and \mathbf{H} is:

$$\mathbf{H} = \begin{pmatrix} \mathbf{L}_\mathbf{u}^1 & 0 & \dots & 0 & \frac{\partial \mathbf{u}^1}{\partial \gamma} \\ 0 & \mathbf{L}_\mathbf{u}^2 & \dots & 0 & \frac{\partial \mathbf{u}^2}{\partial \gamma} \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & \dots & \dots & \mathbf{L}_\mathbf{u}^p & \frac{\partial \mathbf{u}^p}{\partial \gamma} \end{pmatrix} \quad (6)$$

where

$$\frac{\partial \mathbf{u}}{\partial \gamma} = \begin{pmatrix} x & 0 & 1 & 0 & -\frac{p_x \rho x}{Z+\xi\rho} \\ 0 & y & 0 & 1 & -\frac{p_y \rho y}{Z+\xi\rho} \end{pmatrix} \quad \text{and} \quad \mathbf{L}_\mathbf{u}^i = \frac{\partial \mathbf{u}^i}{\partial \mathbf{r}^i} \quad (7)$$

with \mathbf{r}^i , the pose vector linked to image i .

Poses are then updated using the exponential map of $se(3)$ [15] using ${}^c\mathbf{M}_o^{t+1} = {}^c\mathbf{M}_o^t e^{[v]}$ and intrinsic parameters are updated by $\gamma^{t+1} = \gamma^t + \dot{\gamma}$.

III. CENTRAL STEREOSCOPIC CALIBRATION

Extending the mono camera approach, the optimization criterion becomes [13]:

$$\Delta_s = \sum_{j=1}^N \sum_{i=1}^{k_j} (pr_{\gamma_j}({}^{c_j}\mathbf{M}_{c_1} {}^c\mathbf{M}_o, {}^o\mathbf{X}_i) - {}^{c_j}\mathbf{u}_i^*)^2. \quad (8)$$

Considering \mathbf{L}_j , the camera j pose interaction matrix, the stereo pose interaction matrix is then [13]:

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_1 \\ \mathbf{L}_2 {}^{c_2}\mathbf{V}_{c_1} \\ \vdots \\ \mathbf{L}_N {}^{c_N}\mathbf{V}_{c_1} \end{bmatrix} \quad \text{with} \quad {}^{c_j}\mathbf{V}_{c_1} = \begin{bmatrix} {}^{c_j}\mathbf{R}_{c_1} & [{}^{c_j}\mathbf{t}_{c_1}]_\times \\ 0 & {}^{c_j}\mathbf{R}_{c_1} \end{bmatrix}. \quad (9)$$

${}^{c_j}\mathbf{V}_{c_1}$ is the twist transformation matrix between velocity vectors of camera 1 ${}^{c_1}\mathbf{v}$ and camera j ${}^{c_j}\mathbf{v}$. ${}^{c_j}\mathbf{R}_{c_1}$ and ${}^{c_j}\mathbf{t}_{c_1}$ are the rotation and translation blocs of ${}^{c_j}\mathbf{M}_{c_1}$ and $[{}^{c_j}\mathbf{t}_{c_1}]_\times$, the skew-symmetric matrix of ${}^{c_j}\mathbf{t}_{c_1}$. So, with ${}^{c_j}\dot{\mathbf{u}}_i$, the set of time variation of camera j features in image i , the stereoscopic calibration control law for p images is:

$$\begin{pmatrix} \mathbf{v}_1^1 & \dots & \mathbf{v}_1^p & \mathbf{v}_{1,2} & \dots & \mathbf{v}_{1,N} & \dot{\gamma}_1 & \dots & \dot{\gamma}_N \end{pmatrix}^T = \mathbf{H}_s^+ \begin{pmatrix} {}^{c_1}\dot{\mathbf{u}}_1 & \dots & {}^{c_N}\dot{\mathbf{u}}_1 & {}^{c_1}\dot{\mathbf{u}}_2 & \dots & {}^{c_N}\dot{\mathbf{u}}_2 & \dots & {}^{c_N}\dot{\mathbf{u}}_p \end{pmatrix}^T \quad (10)$$

where \mathbf{v}_1^i is the stereo rig pose velocity vector, for each set of N images and $\mathbf{v}_{1,j}$, the relative pose velocity vector between camera j and camera 1. Considering \mathbf{L}^i , the interaction matrix of the stereo rig pose i and $\mathbf{L}_{1,j}^i$, the interaction matrix of relative pose between camera 1 and camera j for pose i , \mathbf{H}_s is expressed in equation (11).

IV. RESULTS

Non-linear optimization methods need an initial guess. Each camera parameters set is coarsely initialized with u_0, v_0 at the image center, p_x, p_y at the half vertical size of the image (mirror radius). ξ is initialized with its known theoretical values [5]. Then, the pose of each camera is initialized with the adaptation we did of a linear pose computation method [16]. The calibration grid is a 82 cm \times 105 cm chessboard of 7×9 squares.

$$\mathbf{H}_S = \left(\begin{array}{ccc|ccc|ccc} 0 & \dots & 0 & 0 & \dots & 0 & \frac{\partial {}^{c_1}\mathbf{u}_1}{\partial \gamma_1} & 0 & \dots & 0 \\ \mathbf{L}^1 & \vdots & \dots & \vdots & \mathbf{L}_{1,2}^1 & \dots & 0 & 0 & \frac{\partial {}^{c_2}\mathbf{u}_1}{\partial \gamma_2} & \dots & 0 \\ & \vdots & \dots & \vdots & \vdots & \ddots & \vdots & \vdots & \dots & \ddots & \vdots \\ & 0 & \dots & 0 & 0 & \dots & \mathbf{L}_{1,N}^1 & 0 & \dots & 0 & \frac{\partial {}^{c_N}\mathbf{u}_1}{\partial \gamma_N} \\ \hline \vdots & \dots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline 0 & \dots & 0 & 0 & \dots & 0 & \frac{\partial {}^{c_1}\mathbf{u}_p}{\partial \gamma_1} & 0 & \dots & 0 \\ \vdots & \dots & 0 & \mathbf{L}^p & \mathbf{L}_{1,2}^p & \dots & 0 & 0 & \frac{\partial {}^{c_2}\mathbf{u}_p}{\partial \gamma_2} & \dots & 0 \\ & \vdots & \vdots & & \vdots & \ddots & \vdots & \vdots & \dots & \ddots & \vdots \\ & 0 & \dots & 0 & 0 & \dots & \mathbf{L}_{1,N}^p & 0 & \dots & 0 & \frac{\partial {}^{c_N}\mathbf{u}_p}{\partial \gamma_N} \end{array} \right). \quad (11)$$

A. Monocular calibration and comparison

In this section, calibration results are evaluated for a single camera modelled by a sphere using the approach presented in section II. Comparison is made with the Mei's toolbox [8].

This experiment show the calibration of a catadioptric camera composed by a camera (Sony DFW-SX910) and a Remote Reality paraboloid optic (one of the two camera visible in figure 1(a)). Six calibration images of 1280×960 pixels resolution are used for a total of 288 points (Fig. 3).



Fig. 3. Monocular image of a calibration chessboard.

Distortions are not calibrated and results are reported in table I. For a fair results comparison, the same set of subpixelic image points, extracted from the corners of the chessboard target, is used.

TABLE I

CALIBRATION RESULTS AND COMPARISON WITH [8]. $\mu_{\mathbf{u}}$ IS THE PIXELIC MEAN ERROR AND $\sigma_{\mathbf{u}}$, THE STANDARD DEVIATION.

| method | VVS | Mei |
|-----------------------|--------|--------|
| p_x | 460.33 | 460.33 |
| p_y | 459.65 | 459.65 |
| u_0 | 635.99 | 635.99 |
| v_0 | 490.19 | 490.19 |
| ξ | 1.14 | 1.14 |
| $\mu_{\mathbf{u}}$ | 0.168 | 0.168 |
| $\sigma_{\mathbf{u}}$ | 0.137 | 0.137 |
| iterations | 21 | 60 |

Calibration results using VVS or Mei's Toolbox are the same, up to 10^{-4} for standard deviation of backprojection error in favour of VVS. This result validates our approach.

B. Stereo calibration

a) *Two omnidirectional cameras calibration:* Calibration results of the two cameras stereo rig (Fig. 1(a)) show an equal estimated value of ξ (Tab. II), which is coherent since mirrors are of the same model and of a precise work. Considering both cameras are vertical, the manually measured horizontal distance between them is about 38.5 cm and the vertical one is 9 cm. Estimation errors of these distances are respectively 0.13 cm and 0.05 cm. Finally, sensors orientations are quasi identical, which is well shown by the weak estimated rotations. Initial values of parameters come from the individual calibration of each catadioptric sensor and the initial relative pose is computed thanks to the chessboard poses in a pair of images.

TABLE II

SIMULTANEOUS INTRINSIC AND EXTRINSIC CALIBRATION OF THE TWO CAMERAS RIG. $\mu_{\mathbf{u}}$ IS THE PIXELIC MEAN ERROR AND $\sigma_{\mathbf{u}}$, THE STANDARD DEVIATION. EXTRINSIC PARAMETERS: {38.33 cm, 1.83 cm, -8.95 cm, 1.33°, -2.02°, -2.20°} WITH A TRANSLATION ERROR OF 0.13 CM.

| camera | c_1 | c_2 |
|-----------------------|--------|--------|
| p_x | 458.51 | 485.14 |
| p_y | 457.01 | 484.91 |
| u_0 | 633.48 | 617.31 |
| v_0 | 498.67 | 527.12 |
| ξ | 1.14 | 1.14 |
| $\mu_{\mathbf{u}}$ | 0.251 | |
| $\sigma_{\mathbf{u}}$ | 0.212 | |

b) *Calibration of the FOO sensor:* With a single orthographic camera and four parabolic mirrors, the FOO sensor is virtually composed by four catadioptric cameras. Calibration results of the FOO sensor are visible in table III. The resolution of images is 2560×1920 pixels (Fig. 4), leading to a similar mirror resolution than for the multiple cameras stereo rig of the

previous experiment. The four mirrors are placed in a square with a side length of 3 cm. The used calibration target is still a chessboard but measuring 20 cm \times 25 cm to be able to place it near the FOO in order to reduce uncertainties due to the weak baseline.

Intrinsic parameters precision is tough to evaluate but if extrinsic parameters are well estimated, it can be deduced that intrinsic ones are correct. The error of extrinsic parameters estimation, *i.e.* ${}^c_j\mathbf{M}_{c_1}$ for $2 \leq j \leq 4$, is computed from the six distances between the four mirrors (Tab. IV). The mean error of the six distances is 1.5 mm, leading to a 4% relative error, computed with the ratio of the mean error over the mean real distance between mirrors.

V. CONCLUSION

A new calibration method of stereoscopic omnidirectional sensor has been presented. The method allows the simultaneous intrinsic and extrinsic calibration of a N cameras rig. Results show the achievement of calibration and are competitive with existing approaches. Finally, this method also allows the calibration of non standard cameras, such as the FOO sensor.

REFERENCES

- [1] S. BAKER, S.K. NAYAR, "A theory of single-viewpoint catadioptric image formation", *Int. Journal of Computer Vision*, Vol. 35, No. 2, november 1999.
- [2] J. GLUCKMAN, S.K. NAYAR, K.J. THORESZ, "Real-time omnidirectional and panoramic stereo", *Proc. of the 1998 DARPA Image Understanding Workshop*, november 1998.
- [3] G. JANG, S. KIM, I. KWEON, "Single Camera Catadioptric Stereo System", *OmniVis, workshop of Int. Conf. on Computer Vision*, Beijing, China, october 2005.
- [4] E. MOUADDIB, R. SAGAWA, T. ECHIGO, Y. YAGI, "Stereo Vision with a Single Camera and Multiple Mirrors", *IEEE Int. Conf. on Robotics and Automation*, Barcelona, Spain, april 2005.
- [5] J.P. BARRETO, H. ARAUJO, "Issues on the Geometry of Central Catadioptric Image Formation", *Int. Conf. On Pattern Recognition*, Tampa, USA, december 2001.
- [6] E. MARCHAND, F. CHAUMETTE, "A New Formulation for Non-Linear Camera Calibration Using Virtual Visual Servoing", *INRIA Research Report, no 4096*, january 2001.
- [7] J.-Y. BOUGUET, "Camera Calibration Toolbox for Matlab", www.vision.caltech.edu/bouguetj.
- [8] C. MEI, P. RIVES, "Single View Point Omnidirectional Camera Calibration from Planar Grids", *IEEE Int. Conf. on Robotics and Automation*, Roma, Italy, april 2007.
- [9] J.P. BARRETO, H. ARAUJO, "Paracatadioptric Camera Calibration Using Lines", *IEEE Int. Conf. on Computer Vision*, Nice, France, october 2003.
- [10] D. SCARAMUZZA, R. SIEGWART "A New Method and Toolbox for Easily Calibrating Omnidirectional Cameras", *Workshop on Camera Calibration Methods for Computer Vision Systems*, Bielefeld, Germany, march 2007.
- [11] V. SUNDARESWARAN, R. BEHRINGER, "Visual servoing-based augmented reality", *IEEE Int. Workshop on Augmented Reality*, San Francisco, USA, november 1998.
- [12] E. MARCHAND, F. CHAUMETTE, "Virtual visual servoing: a framework for real-time augmented reality", *EUROGRAPHICS*, Saarbrücken, Germany, september 2002.
- [13] G. CARON, E. MARCHAND, E. MOUADDIB, "3D Model Based Pose Estimation For Omnidirectional Stereovision", *Int. Conf. on Intelligent Robots and Systems*, Saint-Louis, USA, october 2009.
- [14] JOAO P. BARRETO, F. MARTIN, R. HORAUD, "Visual Servoing/Tracking Using Central Catadioptric Images", *Experimental Robotics VIII.*, B. Siciliano and P.Dario (Ed). Springer Verlag, 2003.
- [15] Y. MA, S. SOATTO, J. KOŠECKÁ, S. SASTRY, "An invitation to 3-D vision", *Springer*, 2004.
- [16] L. QUAN, Z. LAN, "Linear N-Point Camera Pose Determination", *IEEE Trans. On Pattern Analysis And Machine Intelligence*, Vol. 21, No. 7, july 1999.

TABLE III

CALIBRATION OF THE OMNIDIRECTIONAL STEREOVISION SENSOR OF MOUADDIB [4]: THE FOO. THIS SENSOR IS CONSIDERED, HERE, AS A RIG OF FOUR PARACATADIOPTRIC CAMERAS. $\mu_u=0.918$ AND $\sigma_u=0.863$. MEAN EXTRINSIC TRANSLATION ERROR: 0.15 CM.

| | c_1 | c_2 | c_3 | c_4 |
|------------|-------|--------|--------|--------|
| p_x | 367.6 | 388.9 | 397.4 | 426.2 |
| p_y | 367.6 | 389.2 | 398.6 | 431.3 |
| u_0 | 818.1 | 1737.9 | 1726.1 | 810.8 |
| v_0 | 534.9 | 539.1 | 1458.9 | 1447.2 |
| ξ | 0.68 | 0.77 | 0.79 | 0.84 |
| t_x | / | -2.93 | -3.04 | -0.28 |
| t_y | / | 0.13 | -3.00 | -2.93 |
| t_z | / | -0.47 | -0.37 | 0.06 |
| θ_x | / | -0.7 | -2.5 | -2.9 |
| θ_y | / | 2.9 | 2.1 | -0.3 |
| θ_z | / | 0.5 | 0.5 | 0.0 |

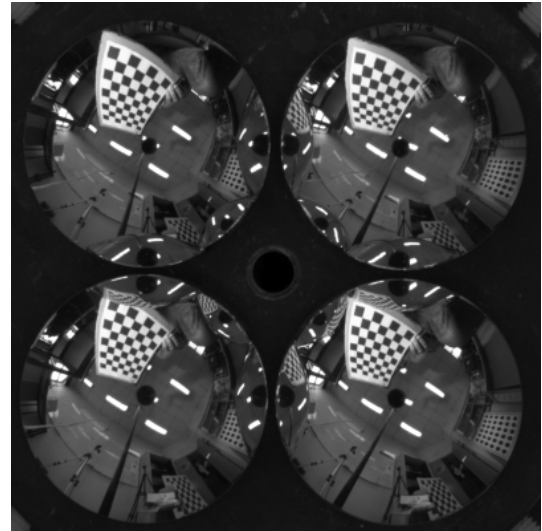


Fig. 4. Stereo-catadioptric image of a calibration chessboard.

TABLE IV

EVALUATION OF EXTRINSIC PARAMETERS ESTIMATION FOR THE FOO SENSOR. $d(c_i, c_j)$ IS THE DISTANCE IN CENTIMETERS BETWEEN MIRRORS i AND j . ERRORS ARE COMPUTED W.R.T. REFERENCE ONES.

| | reference | estimation |
|---------------|-----------|------------|
| $d(c_1, c_2)$ | 3.00 | 2.93 |
| $d(c_2, c_3)$ | 3.00 | 2.87 |
| $d(c_3, c_4)$ | 3.00 | 2.76 |
| $d(c_4, c_1)$ | 3.00 | 2.94 |
| $d(c_1, c_3)$ | 4.24 | 4.27 |
| $d(c_2, c_4)$ | 4.24 | 3.86 |
| mean error | / | 0.15 |